Firefly luciferase gene contains a cryptic promoter

Václav Vopálenský, Tomáa Maaek, OndYej Horváth, et al.

RNA 2008 14: 1720-1729 originally published online August 12, 2008
Access the most recent version at doi:10.1261/rna.831808

Firefly luciferase gene contains a cryptic promoter

VÁCLAV VOPÁLENSKÝ,[1,3] TOMÁŠ MAŠEK,[1,3] ONDŘEJ HORVÁTH,[2] BLANKA VICENOVÁ,[1] MARTIN MOKREJŠ,[1] and MARTIN POSPÍŠEK[1]

[1]Department of Genetics and Microbiology, Faculty of Science, Charles University in Prague, Vinicna 5, 128 44 Prague, Czech Republic
[2]Institute of Molecular Genetics, Laboratory of Leukocyte Antigens, AS CR, Videnska 1083, 140 00 Prague, Czech Republic

ABSTRACT

A firefly luciferase (FLuc) counts among the most popular reporters of present-day molecular and cellular biology. In this study, we report a cryptic promoter activity in the *luc*+ gene, which is the most frequently used version of the firefly luciferase. The FLuc coding region displays cryptic promoter activity both in mammalian and yeast cells. In human CCL13 and Huh7 cells, cryptic transcription from the *luc*+ gene is 10–16 times weaker in comparison to the strong immediate-early cytomegalovirus promoter. Additionally, we discuss a possible impact of the *FLuc* gene cryptic promoter on experimental results especially in some fields of the RNA-oriented research, for example, in analysis of translation initiation or analysis of miRNA/siRNA function. Specifically, we propose how this newly described cryptic promoter activity within the *FLuc* gene might contribute to the previous determination of the strength of the cryptic promoter found in the cDNA corresponding to the hepatitis C virus internal ribosome entry site. Our findings should appeal to the researchers to be more careful when designing firefly luciferase-based assays as well as open the possibility of performing some experiments with the hepatitis C virus internal ribosome entry site, which could not be considered until now.

Keywords: luciferase; cryptic promoter; reporter gene; hepatitis C virus; siRNA; miRNA

INTRODUCTION

A firefly luciferase (FLuc) from the common North American firefly *Photinus pyralis* is very popular nowadays as a very sensitive reporter with an extraordinarily broad dynamic range of the measurable activity. The sensitivity, versatility, and relative simplicity of the luciferase-based assays are the reasons why *FLuc* ranks among the most frequently used reporter genes and as such has been used in thousands of variously designed experiments since the time of its first discovery for application in molecular and cellular biology (de Wet et al. 1987). The broad use of firefly luciferase in research laboratories and for diagnostic purposes was further facilitated by the effort of Promega Corporation that introduced a variety of products designed to simplify and to improve luciferase-based assays. Among these improvements, Promega introduced a version of the firefly luciferase gene (*luc*+) in the pGL3 plasmid series (Groskreutz et al. 1995) that has become the most commonly used version at the present time.

Recently, several researchers have revised the published results by finding cryptic promoters or cryptic splicing sites in the genomic fragments that had previously been claimed to bear an internal ribosome entry sites (for review, see Kozak 2003, 2005; Baird et al. 2006; Mokrejš et al. 2007). Generation of unwanted aberrant transcripts by cryptic splicing sites and/or cryptic promoters present in plasmid backbone and their ability to affect the particular assays also have been reported (Boshart et al. 1992; Rosfjord et al. 1994; Hennecke et al. 2001; Hall et al. 2002; Giannakis et al. 2003; Van Eden et al. 2004; Holcik et al. 2005; Kozak, 2007; Makelainen and Makinen 2007). However, no significant attention has been paid to cryptic transcription from reporter genes until now.

We report here the presence of a cryptic promoter activity in the firefly luciferase gene that is detectable both in yeast and in mammalian cells and that can, due to extremely high sensitivity of the luciferase enzyme assay, seriously affect experimental results. Consistent with this finding, we re-evaluate the published strength of the promoter located in the cDNA sequence corresponding to the internal ribosome entry site of the hepatitis C virus (HCV IRES) (Dumas et al. 2003).

[3]These authors contributed equally to this work.

Reprint requests to: Martin Pospíšek, Department of Genetics and Microbiology, Faculty of Science, Charles University in Prague, Vinicna 5, 128 44 Prague, Czech Republic; e-mail: martin@natur.cuni.cz; fax: 420-2-21951724.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.831808.

RESULTS

The firefly luciferase *luc+* gene displays a cryptic promoter activity both in mammalian and yeast cells

During our studies of translation initiation at the hepatitis C internal ribosome entry site (Mašek et al. 2007) as well as during our theoretical RNA-oriented research (Mokrejš et al. 2006), we gained suspicion that the commonly used *luc+* version of the firefly luciferase gene contains a promoter that is active in various cells of different organisms. To prove this hypothesis, we decided to take advantage of the pRG vector, which was originally developed for the examination of IRES elements by flow cytometry and fluorescence microscopy in our laboratory (Mašek et al. 2007). Schematic drawings of all the mammalian vectors used throughout this study including typical results of the corresponding flow cytometry experiments are shown in Figure 1. The pRG vector (Fig. 1B) contains the red fluorescence protein (DsRED2) gene and the enhanced green fluorescence protein (EGFP) gene as the first and the second reporter cistrons, respectively, and can be used for in vivo production of *DsRED2-EGFP* bicistronic mRNA under the control of the cytomegalovirus immediate-early (*CMV IE*) promoter. The promoterless variant of pRG, the pRG(-P) vector (Fig. 1C), has a reasonably low background of EGFP production and can be used effectively as a tool for detection and analysis of possible cryptic transcription sites located within the inserted intercistronic regions. To test the putative promoter activity of the FLuc protein-coding sequence, we inserted its *luc+* version (Promega) between the *DsRED2* and *EGFP* reporter genes

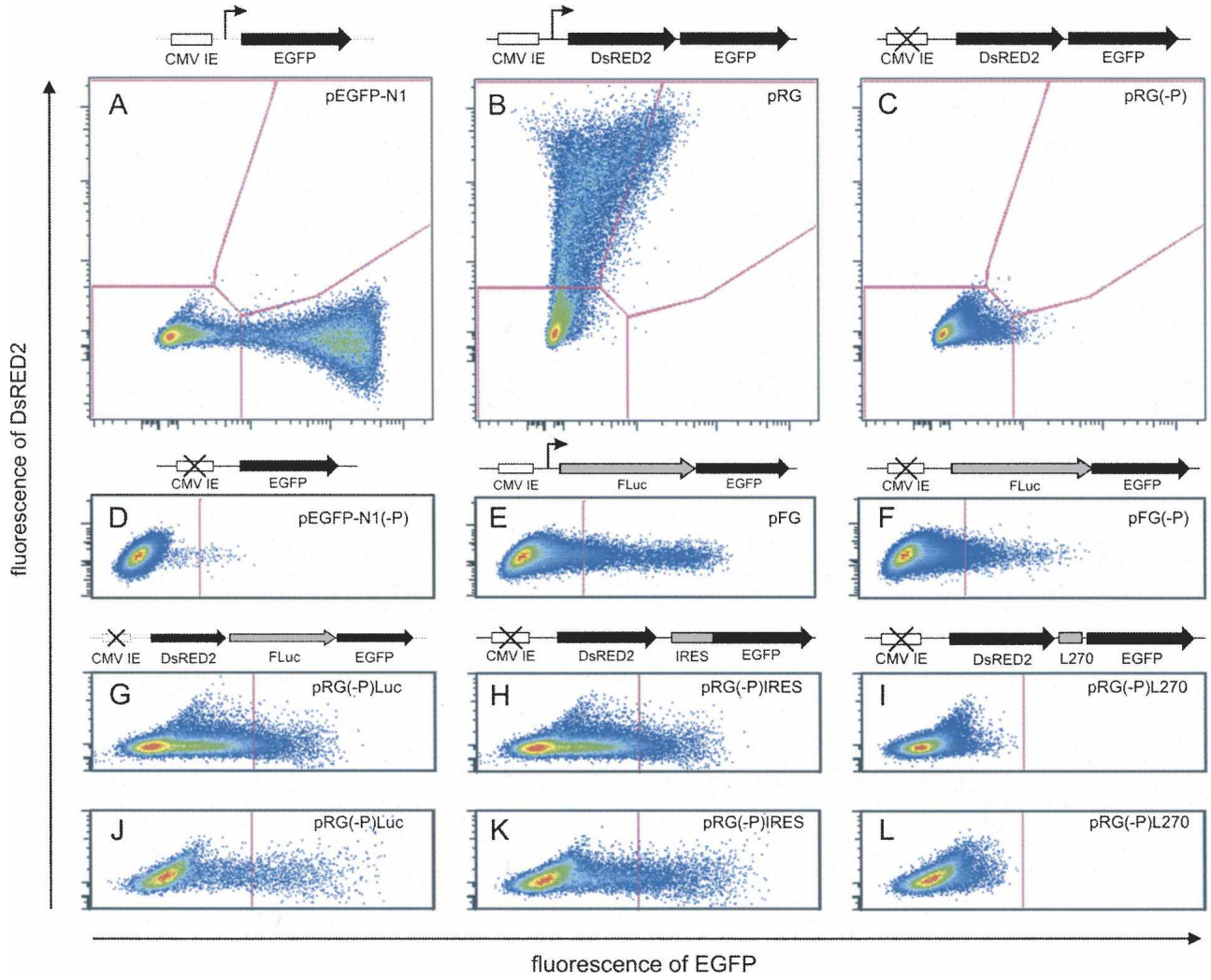

FIGURE 1. Design of the mammalian vectors and corresponding flow cytometry experiments. CCL13 (*A–I*) or Huh7 (*J–L*) cells were transfected with various plasmids with or without promoters. Names of the plasmids are depicted in the *upper-right* corners of the dot plots, corresponding schematic representations of these plasmids are *above* them. Y- and X-axes represent the red and the green fluorescence, respectively. Lines inside the dot plots indicate gating of the corresponding EGFP and DsRED2 cell populations.

of the promoterless backbone, thus creating the pRG(-P)Luc vector (Fig. 1G,J). A small portion of the cells transiently transfected with the negative control promoterless pRG(-P) vector still produced basal levels of EGFP (Fig. 1C). We succeeded in eliminating even this basal production of green fluorescence protein by insertion of the L270 fragment into the intercistronic region. The L270 fragment (IRESite ID 97; http://www.iresite.org) is a short DNA sequence (270 nt) containing a small open reading frame. L270 has been selected from the λ phage DNA library in our previous work for its ability to significantly prevent measurable nonspecific translation of the 3′-cistron from bicistronic mRNA probably by blocking a ribosome readthrough from the 5′ to the 3′ cistron (Mašek et al. 2007). The resulting pRG(-P)L270 vector shows no detectable levels of EGFP production under the experimental conditions described herein and was used to adjust the system baseline of all the measurements (Fig. 1I,L). As a positive control of cryptic promoter activity, we used the cDNA corresponding to the hepatitis C internal ribosome entry site (1 to 385 nt of the HCV 1a genome) (Fig. 1H,K).

In all the flow cytometry experiments, the pEGFP-N1 vector (Clontech) was used both for setting gates for EGFP-producing green fluorescent cells as well as a positive control for measurement of the *CMV IE* promoter strength (Fig. 1A). To gate the signal from the red fluorescence protein, we used the pRG vector itself (Fig. 1B). The two promoterless pRG(-P)-based vectors containing either the *FLuc* gene (Fig. 1G,J) or HCV IRES cDNA (Fig. 1H,K) in the intercistronic region clearly show their cryptic promoter activity, and thus their ability to drive an expression of EGFP in transfected cells. Figure 2A and Table 1 summarize the results of six independent biological replications aimed to determine the strength of the possible promoter activity within both the *luc+* gene and the HCV IRES cDNA. The promoter activity was calculated as a mean of green fluorescence of all the gated EGFP-positive cells. It is clear that both sequences contain the promoter displaying some 16 times lower activity in comparison to the strong immediate-early *CMV* promoter. Relatively high green fluorescence in pRG(-P)-transfected cells (Fig. 2A) is probably a result of some rare events of nonspecifically expressed EGFP that might be caused, for example, by accidental vector reorganization, its integration into the host genome, or by some rare and weak transcription initiation within the vector backbone. This statement is supported by flow cytometry analysis depicted in Figure 1C, which shows just a few cells (0.31% in average of eight

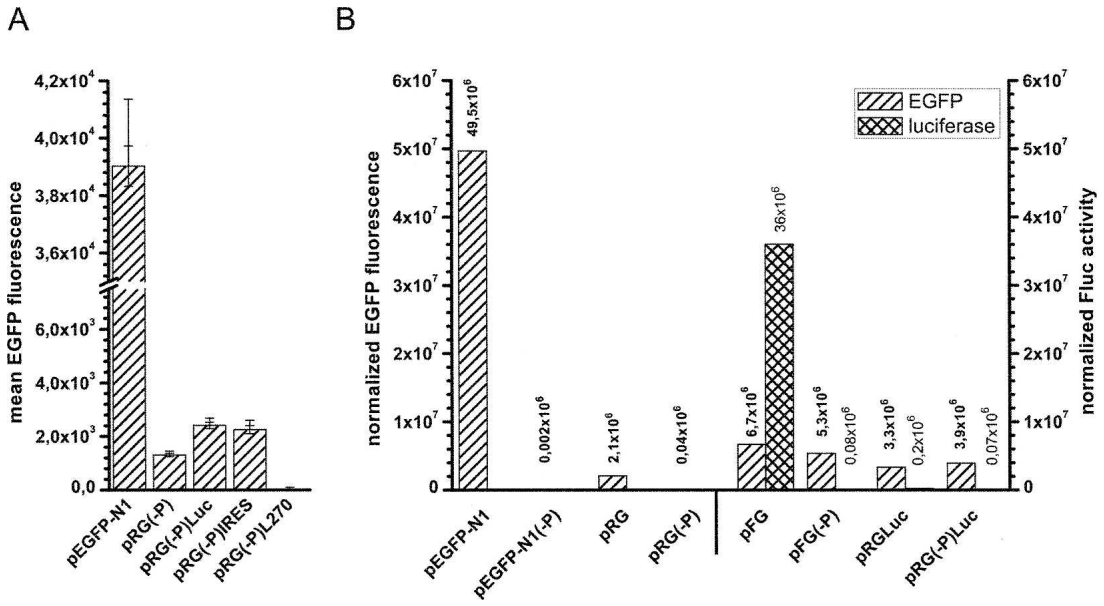

FIGURE 2. Analysis of the cryptic promoter activity associated with the *FLuc* gene and HCV IRES cDNA in mammalian cells. (*A*) Mean EGFP fluorescence per single cell was calculated from the population of positive green fluorescent CCL13 cells expressing EGFP. Both *FLuc* and HCV IRES cDNA display cryptic promoter activity in the promoterless pRG(-P)Luc or pRG(-P)IRES vectors, respectively. Each column represents at least six independent experiments. Unexpectedly, only a few cells (0.31%) of the whole analyzed population appear as green fluorescent also after their transfection with the promoterless pRG(-P) vector. This is probably a result of an aberrant rare transcription [cf. Fig. 1C and FLuc activity of cells transfected with pRGLuc and pRG(-P)Luc in *B*, this figure]. Cells transfected with pRG(-P)L270 show no detectable green fluorescence at all (cf. Fig 1I,L) and were used to set a baseline for these experiments. (*B*) Total EGFP fluorescence and FLuc activity normalized to the number of cells subjected to analysis. Values are expressed as an EGFP fluorescence and/or luciferase activity per 50,000 cells transiently transfected with the pEGFP-N1, pFG, pRG, and pRGLuc vectors and their promoterless variants. The shaded columns and the values in bold *above* them represent total EGFP fluorescence. The cross-shaded columns represent measured FLuc activities. The luciferase activity was only measured in cells transfected with plasmids harboring the *FLuc* gene [pFG, pFG(-P), pRGLuc, pRG(-P)Luc]. The only significant production of the active luciferase was detected in cells transfected with the pFG vector.

TABLE 1. Comparison of the promoter activity of different DNA sequences determined by flow cytometry

Promoter	Vector	EGFP-positive cells mean fluorescence ± SD	Number of biological replications	Promoter strength relative to *CMV IE* (pEGFP-N1) (%)
CMV IE	pEGFP-N1	39026.09 ± 2331.57	11	100
luc+ coding region	pRG(-P)Luc	2411.80 ± 270.58	6	6.2
HCV IRES cDNA	pRG(-P)IRES	2251.60 ± 345.39	6	5.8
Not applicable	pRG(-P)L270	0.02 ± 0.01	6	0

experiments) possessing green fluorescence among the cells transfected with pRG(-P). Furthermore, total EGFP production from the promoterless pRG(-P) control vector normalized to the number of cells subjected to analysis is comparable to background in appearance (Fig. 2B). Compared to that, the positive control vector pEGFP-N1 gives 57.21% of green fluorescent cells in an average of 11 experiments (Fig. 1A). The value of 57.21% of positive green fluorescent cells transfected with the control pEGFP-N1 vector represents also the average transfection efficiency in all the experiments. Throughout the whole study, we used epithelial CCL13 cells, also referred to as Chang cells. However, similar results were obtained also with Huh7 liver cells (Fig. 1J–L) except that the average transfection efficiencies were 44% under the described conditions.

To prove the cryptic promoter activity of the *luc*+ gene in another context, we prepared the new pFG vector, where the *CMV IE* promoter is directly followed by downstream *FLuc* and *EGFP* genes, as well as its promoterless variant pFG(-P), containing only *FLuc* followed by *EGFP*. Typical results obtained by flow cytometry of CCL13 cells transiently transfected with pFG and pFG(-P) vectors are depicted in Figure 1, E and F, and Figure 2B. Contrary to Figure 2A depicting the calculated mean of green fluorescence of every EGFP gated (positive above the set threshold) cell, Figure 2B represents total EGFP fluorescence normalized to 50,000 cells transfected with pEGFP-N1, pRG, pRGLuc, pFG, and their promoterless variants. The results are shown after subtraction of autofluorescence of untransfected CCL13 cells and clearly point to negligible EGFP production from the promoterless pEGFP-N1(-P) vector. This result is consistent with fractionally measurable activity of luciferase in pFG(-P) and pRGLuc (Fig. 2B) and no detectable red fluorescence in pRG(-P) (data not shown) containing cells, and clearly demonstrates no or negligible transcription from plasmid backbones. The significant amount of EGFP produced in cells transfected with promoterless pFG(-P) vector thus gives another strong evidence for the cryptic promoter activity within the *FLuc* gene (Fig. 2B). The strength of the *FLuc* cryptic promoter appears to be only less than 10 times lower than the activity of the strong *CMV IE* promoter in this assay. Assaying luciferase activity in transfected cells also answered the question of whether the shorter transcripts generated from the *FLuc* cryptic promoter can be translated to enzymatically active polypeptides. A significant amount of active luciferase was detected only in cells containing the full-length luciferase gene under the control of the *CMV IE* promoter, while cells transfected with the promoterless pFG(-P) vector displayed almost no luciferase activity (Fig. 2B).

To support the sensitive reporter assays described above with RNA data, we analyzed the presence of transcripts complementary to the FLuc-coding region by Northern blotting and real-time qRT-PCR. The autoradiogram in Figure 3A represents Northern blots of poly(A)$^+$-enriched RNA isolated from CCL13 cells transfected either with pRG or with pRGLuc and/or with their promoterless variants. Membranes were probed with ^{32}P-labeled ssRNA probes against either 5′ or 3′ *FLuc* gene ends, respectively. Both probes produce strong signals against the samples prepared from cells transfected with *CMV* promoter-containing vectors; however, only the probe complementary to the 3′ *FLuc* end (nucleotides 1411 – 1634) is capable of hybridizing with mRNA isolated from cells transfected with the promoterless pRG(-P)Luc vector. We carried out several Northern blot experiments with either total RNA or mRNA samples, but every time obtained diffuse signals instead of expected discrete bands. Mapping of possible cryptic transcription sites within the 1653-bp-long *luc*+ coding region by semiquantitative RT-PCR (data not shown) and real-time qRT-PCR (Fig. 3B,C) surprisingly also exhibits no strictly localized transcription site but, rather, a progressively increasing amount of transcripts in the 5′-to-3′ direction of the *FLuc* gene. Figure 3B demonstrates an example of qRT-PCR analysis in which amplicons of similar length uniformly distributed along the *luc*+ transcript were quantified. The observed difference of five-threshold-cycle (C_t) between the amplicons corresponding to very 5′-end and very 3′-end of the *FLuc* coding region thus reflects a 32 times increase of the relative transcript copy number. Figure 3C shows mapping of the transcription activity borne within the luciferase-coding region in more detail. Two independent real-time qRT-PCR experiments were done with 29 combinations of primers covering the whole *Fluc* and part of the *EGFP* genes of the pRG(-P)Luc vector. The

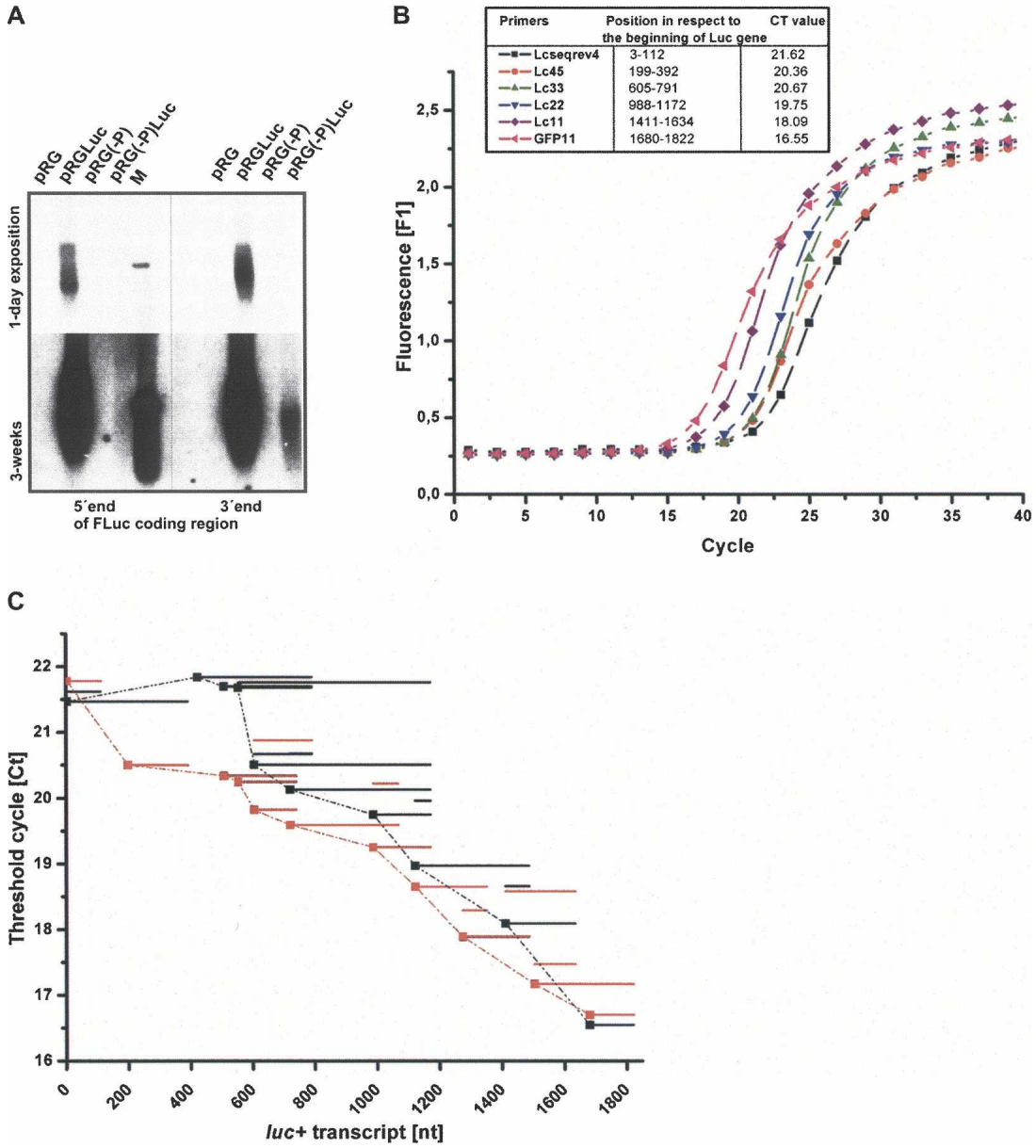

FIGURE 3. Analysis of shorter transcripts generated from the *FLuc* coding region in human CCL13 cells. (*A*) Northern blot analysis of poly(A)$^+$ RNA isolated from CCL13 cells transfected with pRG and pRGLuc vectors and with their promoterless variants (-P), respectively. It is clearly seen that only ssRNA probe complementary to the 3′ half of the *FLuc* coding region produces a signal on the pRG(-P)Luc sample. The obtained signal is in agreement with results depicted in *B* and *C* and probably corresponds to transcription initiation in several sites followed by degradation of aberrant transcripts. Line M is the RNA ladder High Range (Fermentas); the hybridization signal is produced by RNA of 2 kb in length. (*B*) Example of real-time qRT-PCR assay of *FLuc*-related transcripts present in the cells transfected with the promoterless pRG(-P)Luc vector. For this experiment, amplicons of similar length covering the whole *FLuc* coding region step-by-step were selected. The attached table shows positions of the amplicons analyzed as well as corresponding *Ct* values. This result demonstrates a gradual increase of transcripts complementary to *FLuc* CDS from its 5′- to 3′-ends. (*C*) Two independent real-time qRT-PCR experiments (red and black bars, respectively) analyzing production of *FLuc*-related transcripts generated from the pRG(-P)Luc vector in more detail. The length and position of the bars indicate the length and position of the particular amplicon within the *FLuc* transcript. A join connects 5′-ends of the amplicons in each experiment separately. The results show several transcription initiation sites within the *FLuc* region again and correspond to the diffuse banding observable in Northern blot analysis. The results depicted in *B* and *C* point to more than 30 times increase in the abundances of *FLuc* transcripts between 5′- and 3′-ends of the luciferase coding region.

results also display the same trend of gradually increased transcription from 5′- to 3′-ends and are in agreement with Northern blot analysis displaying a diffuse signal rather than distinct bands. The observed gradual increase of C_t values along the *luc+* CDS should not be influenced by possible cryptic transcription from the opposite strand as well as by the presence of RNA degradation species because an *EGFP*-gene-specific primer and poly(A)$^+$ mRNA fraction were

used for cDNA synthesis. We propose that the combination of qRT-PCR and Northern blotting data reflects the real situation in cells, where the transcription from several cryptic promoters scattered along the luciferase-coding region is followed by intensive degradation of aberrant transcripts.

The firefly luciferase is a ubiquitously used reporter gene. Therefore, we tested the possible promoter activity in the protein coding region of its most frequently used variant *luc+* also in one of the most popular eukaryotic model organisms—in the budding yeast *Saccharomyces cerevisiae*. For this purpose we utilized the yeast pFGAL4h vector (Mašek et al. 2007) originally designed for in vivo production of bicistronic mRNAs bearing the *luc+* gene as the first and *GAL4* as the second cistron. To test and measure the possible promoter activity of the *luc+* gene, we deleted the strong *TPI* promoter from the 5′-end of the bicistronic transcription unit and thus prepared a promoterless pFGAL4h(-P) vector (Fig. 4A). Luminescence measurements show a marked drop of the FLuc activity in cells carrying this promoterless pFGAL4h(-P) vector, while the second Gal4 reporter protein remained synthesized, which clearly confirms the activity of the internal *luc+* cryptic promoter in yeast as well (Fig. 4B). To prove the observed cryptic promoter activity by Northern blotting, we prepared a new vector (pYX213Luc) containing the *FLuc* gene under the control of a tightly regulated yeast *GAL1* promoter. Northern blot analyses revealed that the radiolabeled ssRNA probe complementary to the 5′-end of the *FLuc* gene gives a strong signal only with total RNA isolated from yeast harboring the pFGAL4h vector and/or from yeast containing pYX213Luc and growing on galactose. Contrary to that, the probe complementary to the 3′-end of the *FLuc* gene (nucleotides 1411 – 1634) revealed additional shorter transcripts originating from the luciferase-coding region both in yeast transformed with the promoterless variant of the pFGAL4h vector as well as in yeast containing pYX213Luc but growing on glucose, which tightly repressed transcription from the *GAL1* promoter. Measurement of the *FLuc* promoter activity and Northern blotting point to significantly high activity of cryptic transcription starting from several cryptic transcription initiation sites within the *luc+* coding region in yeast.

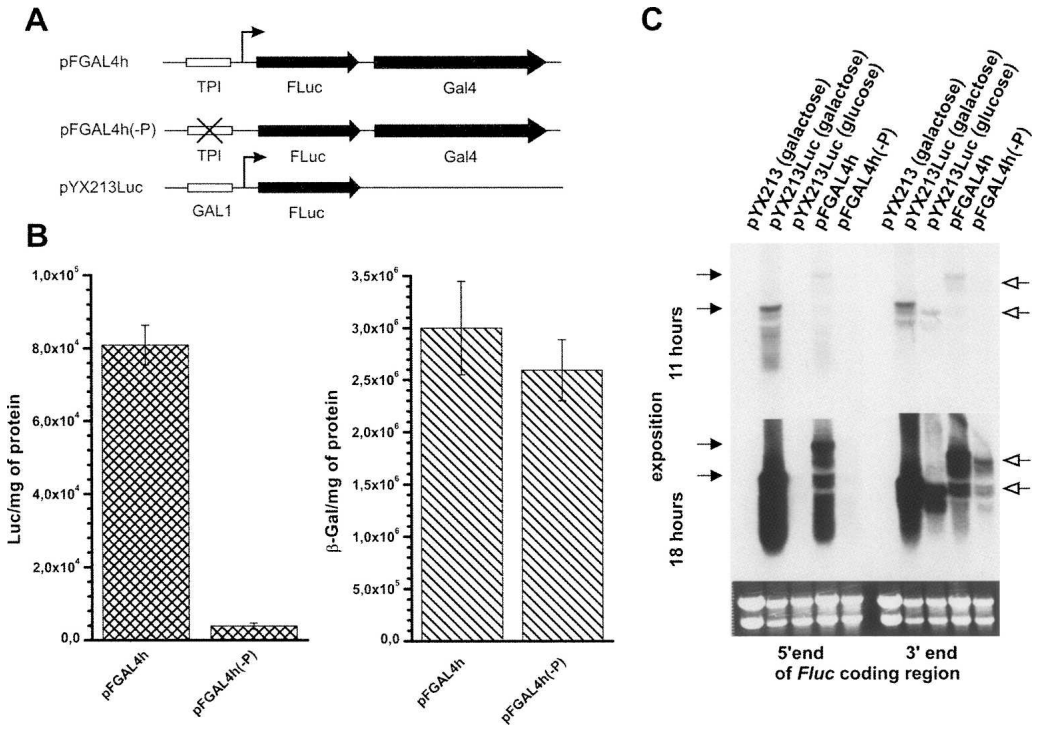

FIGURE 4. The cryptic promoter in the *luc+* gene is active in yeast cells. (*A*) Schematic representation of the yeast vectors used throughout this study. *TPI* is a strong constitutive promoter, and *GAL1* is a strong promoter inducible by galactose. (*B*) Enzymatic activities of FLuc and β-galactosidase normalized to the protein content of cell lysates prepared from yeast strains expressing the following bicistronic vectors: pFGAL4h, empty vector encoding bicistronic mRNA bearing the *luc+* and *Gal4* genes as the first and the second cistrons, respectively; pFGAL4h(-P), the promoterless vector in which the strong constitutive *TPI* promoter was removed. Even limited translation of Gal4p transcriptional activator triggers high expression of β-galactosidase secondary reporter. (*C*) Northern blot analysis of *FLuc* transcripts in yeast strains transformed with vectors depicted in *A*. (Filled arrows) Point to the full-length transcripts generated from pYX213Luc in galactose and from pFGAL4h. (Empty arrows) Point to the most abundant short transcripts generated from the promoterless version of pFGAL4h or from pYX213Luc under glucose repression. (*Bottom*) Amount of total RNA loaded into each line documented by the corresponding electrophoretogram of ribosomal RNAs. Note that the shortened versions of the *FLuc* mRNA hybridize only with the probe directed to the 3′-end of the *FLuc* coding region.

Reevaluation of the cryptic promoter activity in cDNA of the HCV IRES

Michel Ventura's group (Dumas et al. 2003) reported recently that a cDNA fragment corresponding to the hepatitis C virus internal ribosome entry site possesses cryptic promoter activity in human hepatoma cell lines HepG2 and Huh7. We proved this finding in the Huh7 cells and further have shown this activity also in the CCL13 (HeLa) human epithelial cells. However, by careful and exhaustively repeated measurements of the HCV IRES promoter strength, we found it to be much weaker both in CCL13 and Huh7 cell lines (Figs. 1H,K, 2A; Table 1) than previously thought. One explanation of such a discrepancy could be—at least in part of Ventura's experiments—in their promoterless vector design, where *FLuc* as the 5′-reporter directly preceded HCV IRES and the following 3′-reporter (*Renilla* luciferase and/or EGFP). In such constructs, the final level of the 3′-reporter transcription will be a result of combined action of the two cryptic promoters, the HCV IRES promoter and the *FLuc* promoter described herein. Other technical reasons including differences in equipment settings might also contribute to the observed differences in measured strength of the cryptic promoter localized within the HCV IRES cDNA.

DISCUSSION

We present here clear evidence that the *luc+* variant of the firefly luciferase gene contains a cryptic promoter site, which is active both in human cells and yeast. Additionally, both in yeast and human cells, we found a gradual increase of cryptic transcription along the *luc+* coding region reflecting an existence of several transcription initiation sites. In mammalian cells, the observable cryptic transcription from the luciferase-coding region increases total EGFP production (normalized to 50,000 cells) from promoterless *luc+* containing vectors 97.5 times for the pRG(-P)Luc and 2650 times for pFG(-P) when using pRG(-P) and pEGFP-N1(-P) promotor-less vectors as negative controls, respectively (Fig. 2B). From another point of view, the cumulative *FLuc* cryptic promoter is 10 to 16 times weaker than the strong immediate-early promoter from human cytomegalovirus, depending on the vectors used or on the method of statistical evaluation of the obtained data. There is almost no difference between absolute normalized values of EGFP production from pFG(-P) and pRG(-P)Luc vectors (Fig. 2B); thus the higher variance in relative increase of *luc+* promoter activity measured by these two vectors is due to the difference of background EGFP production from their negative control vectors—pEGFP-N1(-P) and pRG(-P), respectively. Deletion of the *CMV IE* promoter from all the vectors almost eliminates translation of the first cistron, suggesting almost no transcription from the pEGFP-N1 plasmid backbone [see Fig. 2B, EGFP fluorescence in pEGFP-N1(-P) and luciferase activity in pFG(-P)]. It thus appears that there probably could be a very weak cryptic transcription coming also from the *DsRED2* reporter. However, quite high production of EGFP from the pRG vector is probably, rather, a result of the ribosome read-through, which becomes clear after its comparison with EGFP production from promoterless pRG(-P) (Fig. 2B). In these two constructs, the *DsRED2* and *EGFP* genes are separated by a 69-nt-long spacer that is short enough to allow measurable read-through (Fig. 3B; Rajkowitsch et al. 2004; Mašek et al. 2007). Such a small difference in EGFP production can be observed only in a well-controlled experiment when determining total normalized EGFP fluorescence after subtraction of empty cells' autofluorescence. Contrary to that, calculating the EGFP mean fluorescence only from cells exhibiting some EGFP signal above threshold gives us, in our opinion, better reproducibility and more precise measurement in higher values, but can be influenced by just several incidentally fluorescent cells, as can be seen in the results of the pRG(-P) vector depicted in Figure 1C and a corresponding column in Figure 2A. The problems with ribosome read-through and possible low cryptic transcription from the *DsRED2* gene in the pRG vector series led us to use the pRG(-P)L270 vector, containing an artificial translation-blocking intercistronic sequence, as one of the controls in our flow cytometry experiments.

Generation of several types of aberrant transcripts also raised a question about their translatability and possible formation of truncated FLuc-EGFP fusion proteins. The luciferase-coding region contains additional 27 AUGs, 53 termination codons, and 11 short open reading frames, suggesting a low probability of production of significantly long polypeptide by efficient translation of shorter transcripts. Similarly, it is very unlikely that higher EGFP expression can be accounted for by translation of some fusion proteins because of different reading phases of *FLuc* and *EGFP* genes in all used plasmids. As it is clearly seen from Figure 2B, the only vector causing appearance of significant FLuc activity in transfected cells is pFG containing *FLuc* as the first cistron and an intact *CMV IE* promoter. The bioinformatical analysis of the *FLuc* coding region together with obtained data showing a gradual increase in transcripts from 5′- to 3′-*FLuc* ends led us to speculate that an increase of EGFP expression reflects only translation of transcripts originating from the end of the luciferase-coding region. This presumption is in agreement with the fact that the last AUG is located in position 1477, which closely precedes the transcription initiation start from the predicted SP1 transcription factor binding site (FunSiteP2.1; http://compel.bionet.nsc.ru/). If this is true, the activity of cryptic transcription from the *luc+* gene in the sense of production of aberrant transcripts is much higher than could be estimated from indirect measurements of protein reporter production.

The firefly luciferase gene is often utilized for preparation of reporter vectors designed for precise and well-controlled transcription of the reporter mRNA in vivo. Typical research areas where the results of luciferase assays are prone to be affected by our findings are those focused on miRNA/siRNA, translation initiation, mRNA polyadenylation, the nonsense-mediated mRNA decay, the 3′-UTR-mediated control of gene expression and mRNA stability, and, of course, those aimed toward determination of promoter strength.

An example of possible influence of cryptic promoter activity within the *FLuc* region on the measured data provided herein is the measurement of cryptic promoter strength in HCV IRES cDNA presented by Dumas et al. (2003), and by us herein. The high popularity of the *FLuc* reporter in miRNA/siRNA-oriented research as well as the availability of the commercial systems dedicated to the miRNA analysis led us to hypothesize that the cryptic promoter activity located in the *FLuc* gene can also influence experimental results obtained in this field. Function and activity of miRNAs and siRNAs are often detected by mRNA containing the FLuc reporter gene and one or multiple target sequences located either in the 3′-untranslated region or between the two reporter cistrons of the bicistronic mRNA or even directly within the FLuc coding region (e.g., Boutla et al. 2003; Yi et al. 2003; Zeng and Cullen 2003; Zeng et al. 2003; Cheng et al. 2005; Rehwinkel et al. 2005; Wu et al. 2005; Martin et al. 2006; Lytle et al. 2007; Martin et al. 2007). Systems for sensitive analysis of microRNA activity based on the *FLuc* gene and insertion of miRNA-binding sites into its 3′-UTR are also offered by several companies including pMIR-REPORT from Ambion. We propose that in such an experimental setup, the presence of the promoter within the *FLuc* coding sequence may result in additional production of shorter mRNAs that do not code for the functional luciferase enzyme but still function as miRNA/siRNA targets and can titrate out some portion of the examined miRNA/siRNA. Under the usual experimental circumstances when the system is oversaturated with tested mRNAs and miRNAs/siRNAs, production of such shorter mRNAs might not change the qualitative answer from the experimental data but certainly will affect the absolute values. Care should be especially taken when the experiment is carried out under more natural conditions either when low levels of tester mRNA production in vivo or low amounts of miRNA/siRNA for direct transfection are used (see, e.g., Boutla et al. 2003). In our opinion, direct evidence of the possible influence of the *FLuc* cryptic promoter on the experimental data in miRNA research came out recently through the work of Lytle et al. (2007), who studied the effect of positional changes of *let-7a* miRNA-binding sites in an mRNA target on its translational repression. In an attempt not to oversaturate the system, they "took care to use the lowest possible amount of DNA or RNA in each transfection." They found that the presence of miRNA-binding sites led to translational repression regardless of whether or not they are located in the 5′-UTR or the 3′-UTR of the *FLuc*-containing monocistronic mRNA—with one exception. They found no repression in the case when the *FLuc* reporter mRNA contained the miRNA-binding site in the 3′-UTR and was produced in vivo after DNA plasmid transfection. Curiously, repression of the similarly designed mRNA was obtained after direct RNA transfection into the cell. Furthermore, the repression also reappeared when the plasmid DNA was electroporated into the cell instead of using common liposome-mediated transfection. Liposome-mediated transfection is known to induce an expression of a number of the cellular genes (Calvin et al. 2006). We therefore assume that the cationic lipid transfection protocol used by Lytle et al. (2007) influenced the analysis of miRNA function by enhancement of cryptic promoter activity within the *FLuc* coding region.

Our findings as well as other reports about the unexpected transcription from cryptic or unusual sites are consistent with the current view on the eukaryotic transcription when most of the genome is transcribed and many parts of it can serve as unconventional transcription starts (ENCODE Project Consortium et al. 2007; Greally 2007). Experiments carried out in yeast are especially prone to be affected by cryptic transcription because *Saccharomyces cerevisiae* can use very short and simple sequences to initiate transcription from both regular and unconventional promoters (Robinson and Lopes 2000; Hellen and Sarnow 2001; Hecht et al. 2002). Such findings should also appeal to the researchers to consider the possibility of cryptic transcripts generated from plasmid backbones and reporter genes in all cases when possible aberrant transcripts may have a potential to influence the results of the planned experiments.

MATERIALS AND METHODS

Plasmid constructs

The pEGFP-N1 vector was obtained from Clontech. The pRG vector is a bicistronic derivative of pDsRed2-C1 (Clontech) containing the *DsRED2* gene and *EGFP* gene as the first and the second cistrons, respectively, under the control of the cytomegalovirus immediate-early promoter. Construction of pRG and pRGL270 vectors were described previously (Mašek et al. 2007). Promoterless vectors [pRG(-P) and pFG(-P)] and their derivatives were prepared by excising the VspI/NheI fragment containing the *CMV IE* promoter, filling in the recessed 3′-termini by the Klenow fragment of the DNA polymerase I and by subsequent re-ligation of blunt-ended vector. The *Photinus* (firefly) luciferase cassette was amplified from the pGL3-Basic plasmid (Promega) with the forward primer (5′-GCGTCGACCATGGAAGACGCCAAAAAC-3′) and the reverse primer (5′-ACGGATCCTTACACGGCG ATCTTTCCG-3′) containing SalI and BamHI restriction sites,

respectively. The whole-length HCV IRES sequence from the HCV1a genotype (bases 1– 385) was amplified from the vector p90HVCFL (courtesy of Charles Rice) with the forward primer (5′-AAAGTCGACGCCAGCCCCCTGATGGGGGCGACAC-3′) and the reverse primer (5′-ACGGATCCGTGTTACGTTTGGTTT TTCTTTGAGGTTTAGG-3′) again containing SalI and BamHI restriction sites, respectively. Both cassettes were inserted into the SalI and BamHI sites of the pRG(-P) vector to create pRG(-P)Luc and pRG(-P)IRES vectors, respectively. To create pRGLuc and pFG vectors, the firefly luciferase cassette was excised from the pRG(-P) vector using SalI and BamHI restriction endonucleases and inserted into the SalI and BamHI sites of pRG and pEGFP-N1 vectors (Clontech), respectively. The yeast pFGAL4h vector was described previously (Mašek et al. 2007). To create its promoterless variant pFGAL4h(-P), the triose isomerase promoter (*TPI*) sequence was removed by AatII and NcoI, and the vector was circularized by ligation with oligonucleotide cassette 5′-CAATTA ACCCTCACTAAAGGGAAAGATCTC-3′ containing a unique BglII site. The insertion of the *FLuc* coding region under the control of the galactose-inducible *GAL1* promoter was carried out in the pYX213 vector (Ingenius) using NcoI and AvrII restriction enzymes. All clones were verified by restriction endonuclease digestion and sequencing.

Cell cultures, DNA transfection, and flow cytometry analysis

The human epithelial cell line CCL13 (also known as Chang cells) and human hepatoma cell line Huh7 were cultured in Dulbecco's modified Eagle's medium (DMEM; Sigma) supplemented with 2 mM L-glutamine, 100 iu/mL penicillin, 100 μg/mL streptomycin, and 10% fetal bovine serum (GIBCO) at 37°C in 5% CO_2 and 95% relative humidity. For transient transfections, cells were plated in 24-well tissue culture plates 24 h before transfection. For the standard flow cytometry analysis, cells were transiently transfected under conditions recommended by the manufacturer by the mixture of 1 μg of DNA in 100 μL of 100 mM NaCl and 3.9 μL of ExGen transfection reagent (Fermentas) per well. Forty-eight hours after transfection, the cells were collected by trypsinization and resuspended in DMEM to 10^6 cells per mL. Samples were analyzed by flow cytometry using a BD LSRII device and a Coherent Sapphire 488-20 DPSS laser to excite cells at 488 nm, a 530/30 nm bandpath filter to detect EGFP, and a 585/42 nm bandpath filter to detect DsRED2 expression. For RNA analysis, the CCL13 cells were seeded 24 h before transfection in a T-75 flask. Transfection was carried out according to the manufacturer's guidelines. The plasmid DNA (40 μg) was diluted in 150 mM sodium chloride to a final volume of 800 μL, vortexed, centrifuged, and mixed with 7 equivalents (103 μL) of ExGen transfection reagent (Fermentas). Forty-eight hours after transfection, the cells were washed once with 10 mL of PBS and subjected to lysis as described further.

The PJ69-4A *S. cerevisiae* strain (MATa, *trp1-901*, *leu2-3,112*, *ura3-52*, *his3-200*, *gal4*Δ, *gal80*Δ, *GAL2-ADE2*, *LYS2*::*GAL1-HIS3*, *met2*::*GAL7-lacZ*) (James et al. 1996) was used for all analyses of transcriptional activity with the pFGAL4h vector series. Experiments with derivatives of pYX213 vectors were carried out in the standard W303-1a *S. cerevisiae* strain (MATa, *leu2-3,112*, *trp1-1*, *can1-100*, *ura3-1*, *ade2-1*, *his3-11,15*).

RNA isolation, Northern blot, and real-time qRT PCR analysis

Human cells were disrupted by an RNA–DNA stabilization reagent (Roche). The mRNA Isolation Kit for Blood and Bone Marrow (Roche) was used to isolate poly(A)$^+$ mRNA. DNase treatment and inactivation were carried out by the DNA-Free kit (Ambion) following the manufacturer's instructions. Yeast total RNA was isolated by TRI reagent (Sigma) and acid-washed glass beads in order to improve the cell disruption. Either 500 ng of mRNA or 15 μg of total RNA were run on a 1% agarose gel according the protocol described in Mašek et al. (2005). ^{32}P-labeled antisense ssRNA hybridization probes displaying complementarity to *luc*+ CDS in regions 3–392 and 1411–1634 bp were synthesized using the T7 RNA transcription kit (Fermentas). Hybridization was carried out in Dig Easy Hyb solution according to Roche's instructions. Poly(A)$^+$ mRNA isolated from 8×10^4 cells and an EGFP gene-specific primer (5′-GCCGTAGGTCAGG GTGGT-3′) were used for cDNA synthesis with SuperScript II RNaseH$^-$ reverse transcriptase (Invitrogen). Two microliters of reverse transcriptase reaction were subjected to PCR amplification (15 min at 95°C; then 40 cycles of 30 sec at 94°C; 30 sec at 54°C; 1 min at 68°C; and finally, 4 min at 72°C) using a LightCycler 1.5 (Roche) and QuantiTect SYBR Green PCR kit (QIAGEN). Control reactions without reverse transcriptase were used in parallel to check that there was no significant DNA contamination in the mRNA samples.

Luciferase and β-galactosidase assays

The Gal4 reporter used throughout this study is a specific transcription factor that activates transcription of the particular secondary reporters in specially engineered yeast strains. The complete description of this reporter system as well as methods of assaying both firefly luciferase and β-galactosidase activities with respective luminescent substrates were described previously (Mašek et al. 2007). Determination of protein concentration in yeast cell lysates was carried out in triplicate using the Bradford reagent (Sigma) and the microplate reader SpectraMax 340PC (Molecular Devices) following the manufacturers' recommendations. Mammalian cells lysates were prepared by adding the Passive lysis buffer (Promega) to the pelleted cells. Firefly luciferase activity was quantified in quadruplicate with the Luciferase Assay System (Promega) according to the manufacturer's instructions. Microlite TLX2 Dynatech Luminometer was used for all luminescence measurements.

ACKNOWLEDGMENTS

We thank Jitka Forstova, Vlasta Pelechova (Charles University, Prague, Czech Republic), and Karel Drbal (Institute of Molecular Genetics, Prague, Czech Republic) for their help. This work was supported by the Czech Science Foundation (grants no. 204/03/1487 and 301/07/0607), by the Grant Agency of Charles University (grant no. 251/2004/B-BIO/PrF), and by the Ministry of Education, Youth and Sports of the Czech Republic (MSM0021620813, MSM0021620858, LC06066).

Received September 19, 2007; accepted June 5, 2008.

REFERENCES

Baird, S.D., Turcotte, M., Korneluk, R.G., and Holcik, M. 2006. Searching for IRES. *RNA* **12:** 1755–1785.

Boshart, M., Kluppel, M., Schmidt, A., Schutz, G., and Luckow, B. 1992. Reporter constructs with low background activity utilizing the cat gene. *Gene* **110:** 129–130.

Boutla, A., Delidakis, C., and Tabler, M. 2003. Developmental defects by antisense-mediated inactivation of micro-RNAs 2 and 13 in *Drosophila* and the identification of putative target genes. *Nucleic Acids Res.* **31:** 4973–4980.

Calvin, S., Wang, J., Emch, J., Pitz, S., and Jacobsen, L. 2006. FuGENE® HD transfection reagent: Choice of a transfection reagent with minimal off-target effect as analyzed by microarray transcriptional profiling. *Biochemica* **4:** 22–25.

Cheng, A.M., Byrom, M.W., Shelton, J., and Ford, L.P. 2005. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res.* **33:** 1290–1297.

de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R., and Subramani, S. 1987. Firefly luciferase gene: Structure and expression in mammalian cells. *Mol. Cell. Biol.* **7:** 725–737.

Dumas, E., Staedel, C., Colombat, M., Reigadas, S., Chabas, S., Astier-Gin, T., Cahour, A., Litvak, S., and Ventura, M. 2003. A promoter activity is present in the DNA sequence corresponding to the hepatitis C virus 5′ UTR. *Nucleic Acids Res.* **31:** 1275–1281.

ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816.

Giannakis, G., Edmondson, S.R., Favaloro, J.M., Zajac, J.D., and Greenland, K.J. 2003. Aberrant cryptic responsiveness of the pCAT 3- and pGL3-promoter reporter vectors. *Biotechniques* **35:** 332–339.

Greally, J.M. 2007. Genomics: Encyclopaedia of humble DNA. *Nature* **447:** 782–783.

Groskreutz, D.J., Sherf, B.A., Wood, K.V., and Schenborn, E.T. 1995. Increased expression and convenience with the new pGL3 luciferase reporter vectors. *Promega Notes* **50:** 2–6.

Hall, M.C., Young, D.A., Rowan, A.D., Edwards, D.R., and Clark, I.M. 2002. Cryptic promoter activity of pBLCAT3 induced by overexpression of AP1 factors. *Biotechniques* **33:** 1004–1008.

Hecht, K., Bailey, J.E., and Minas, W. 2002. Polycistronic gene expression in yeast versus cryptic promoter elements. *FEMS Yeast Res.* **2:** 215–224.

Hellen, C.U. and Sarnow, P. 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes & Dev.* **15:** 1593–1612.

Hennecke, M., Kwissa, M., Metzger, K., Oumard, A., Kroger, A., Schirmbeck, R., Reimann, J., and Hauser, H. 2001. Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs. *Nucleic Acids Res.* **29:** 3327–3334.

Holcik, M., Graber, T., Lewis, S.M., Lefebvre, C.A., Lacasse, E., and Baird, S. 2005. Spurious splicing within the XIAP 5′ UTR occurs in the Rluc/Fluc but not the βgal/CAT bicistronic reporter system. *RNA* **11:** 1605–1609.

James, P., Halladay, J., and Craig, E.A. 1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144:** 1425–1436.

Kozak, M. 2003. Alternative ways to think about mRNA sequences and proteins that appear to promote internal initiation of translation. *Gene* **318:** 1–23.

Kozak, M. 2005. A second look at cellular mRNA sequences said to function as internal ribosome entry sites. *Nucleic Acids Res.* **33:** 6593–6602.

Kozak, M. 2007. Lessons (not) learned from mistakes about translation. *Gene* **403:** 194–203.

Lytle, J.R., Yario, T.A., and Steitz, J.A. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc. Natl. Acad. Sci.* **104:** 9667–9672.

Makelainen, K.J. and Makinen, K. 2007. Testing of internal translation initiation via dicistronic constructs in yeast is complicated by production of extraneous transcripts. *Gene* **391:** 275–284.

Martin, M.M., Lee, E.J., Buckenberger, J.A., Schmittgen, T.D., and Elton, T.S. 2006. MicroRNA-155 regulates human angiotensin II type 1 receptor expression in fibroblasts. *J. Biol. Chem.* **281:** 18277–18284.

Martin, M.M., Buckenberger, J.A., Jiang, J., Malana, G.E., Nuovo, G.J., Chotani, M., Feldman, D.S., Schmittgen, T.D., and Elton, T.S. 2007. The human angiotensin II type 1 receptor +1166 A/C polymorphism attenuates microRNA-155 binding. *J. Biol. Chem.* **282:** 24262–24269.

Mašek, T., Vopálenský, V., Suchomelová, P., and Pospíšek, M. 2005. Denaturing RNA electrophoresis in TAE agarose gels. *Anal. Biochem.* **336:** 46–50.

Mašek, T., Vopálenský, V., Horváth, O., Vortelová, L., Feketová, Z., and Pospíšek, M. 2007. Hepatitis C virus internal ribosome entry site initiates protein synthesis at the authentic initiation codon in yeast. *J. Gen. Virol.* **88:** 1992–2002.

Mokrejš, M., Vopálenský, V., Kolenatý, O., Mašek, T., Feketová, Z., Sekyrovà, P., Skaloudová, B., Kriz, V., and Pospíšek, M. 2006. IRESite: The database of experimentally verified IRES structures (www.iresite.org). *Nucleic Acids Res.* **34:** D125–D130.

Mokrejš, M., Vopálenský, V., Mašek, T., and Pospíšek, M. 2007. Bioinformatical approach to the analysis of viral and cellular internal ribosome entry sites. In *New messenger RNA research communications* (ed. L.B. Kwang), pp. 133–166. Nova Science, Hauppauge, NY.

Rajkowitsch, L., Vilela, C., Berthelot, K., Ramirez, C.V., and McCarthy, J.E. 2004. Reinitiation and recycling are distinct processes occurring downstream of translation termination in yeast. *J. Mol. Biol.* **335:** 71–85.

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* **11:** 1640–1647.

Robinson, K.A. and Lopes, J.M. 2000. The promoter of the yeast INO4 regulatory gene: A model of the simplest yeast promoter. *J. Bacteriol.* **182:** 2746–2752.

Rosfjord, E., Lamb, K., and Rizzino, A. 1994. Cryptic promoter activity within the backbone of a plasmid commonly used to prepare promoter/reporter gene constructs. *In Vitro Cell. Dev. Biol. Anim.* **30A:** 477–481.

Van Eden, M.E., Byrd, M.P., Sherrill, K.W., and Lloyd, R.E. 2004. Demonstrating internal ribosome entry sites in eukaryotic mRNAs using stringent RNA test procedures. *RNA* **10:** 720–730.

Wu, M.T., Wu, R.H., Hung, C.F., Cheng, T.L., Tsai, W.H., and Chang, W.T. 2005. Simple and efficient DNA vector-based RNAi systems in mammalian cells. *Biochem. Biophys. Res. Commun.* **330:** 53–59.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Dev.* **17:** 3011–3016.

Zeng, Y. and Cullen, B.R. 2003. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9:** 112–123.

Zeng, Y., Yi, R., and Cullen, B.R. 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl. Acad. Sci.* **100:** 9779–9784.